Article No. jmbi.1999.3199 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 293, 835–853

JMB

DNA Structure and Polymerase Fidelity

Youri Timsit

Institut de Biologie Physico-Chimique, CNRS - UPR 9080 13, rue Pierre et Marie Curie 75005, Paris, France

The accuracy of DNA replication results from both the intrinsic DNA polymerase fidelity and the DNA sequence. Although the recent structural studies on polymerases have brought new insights on polymerase fidelity, the role of DNA sequence and structure is less well understood. Here, the analysis of the crystal structures of hotspots for polymerase slippage including $(CA)_n$ and $(A)_n$ tracts in different intermolecular contexts reveals that, in the *B*-form, these sequences share common structural alterations which may explain the high rate of replication errors. In particular, a two-faced "Janus-like" structure with shifted base-pairs in the major groove but an apparent normal geometry in the minor groove constitutes a molecular decoy specifically suitable to mislead the polymerases. A model of the rat polymerase β bound to this structure suggests that an altered conformation of the nascent template-primer duplex can interfere with correct nucleotide incorporation by affecting the geometry of the active site and breaking the rules of base-pairing, while at the same time escaping enzymatic mechanisms of error discrimination which scan for the correct geometry of the minor groove.

In contrast, by showing that the *A*-form greatly attenuates the sequence-dependent structural alterations in hotspots, this study suggests that the *A*-conformation of the nascent template-primer duplex at the vicinity of the polymerase active site will contribute to fidelity. The *A*-form may play the role of a structural buffer which preserves the correct geometry of the active site for all sequences. The detailed comparison of the conformation of the nascent template-primer duplex in the available crystal structures of DNA polymerase-DNA complexes shows that polymerase β, the least accurate enzyme, is unique in binding to a *B*-DNA duplex even close to its active site. This model leads to several predictions which are discussed in the light of published experimental data.

Keywords: unusual DNA structure; spontaneous mutagenesis; microsatellite instability

Introduction

The accuracy of DNA replication is fundamental for the genetic stability of the cell. From bacteria to higher eukaryotes, error frequencies are remarkably low, between 10^{-9} and 10^{-10} per base replicated (Echols & Goodman, 1991). These low mutation rates are achieved by multiple steps of error discrimination including base selection by DNA polymerases, 3′-5′ exonucleolytic proofreading and post-replicative DNA repair (Kunkel, 1992). Although the DNA synthesis errors play a role in aging and disease, spontaneous mutations also provide the opportunity for genetic variation and are a primary basis for the evolution. Replication errors occur non-randomly and result from an intricate interplay of intrinsic polymerase fidelity and DNA seq

A principal
the geometric
into DNA (E
1997). Crystal
duplexes cont
shown that t
pseudo 2-fold
base-pair is
providing a geometrical means for selection by the

E-mail address of the corresponding author: timsit@ibpc.fr

Article No. jmbi.1999.3199 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 293, 835–853

JMB

DNA Structure and Polymerase Fidelity

Youri Timsit

Institut de Biologie Physico-Chimique, CNRS - UPR 9080 13, rue Pierre et Marie Curie 75005, Paris, France

The accuracy of DNA replication results from both the intrinsic DNA polymerase fidelity and the DNA sequence. Although the recent structural studies on polymerases have brought new insights on polymerase fidelity, the role of DNA sequence and structure is less well understood. Here, the analysis of the crystal structures of hotspots for polymerase slippage including $(CA)_n$ and $(A)_n$ tracts in different intermolecular contexts reveals that, in the *B*-form, these sequences share common structural alterations which may explain the high rate of replication errors. In particular, a two-faced "Janus-like" structure with shifted base-pairs in the major groove but an apparent normal geometry in the minor groove constitutes a molecular decoy specifically suitable to mislead the polymerases. A model of the rat polymerase β bound to this structure suggests that an altered conformation of the nascent template-primer duplex can interfere with correct nucleotide incorporation by affecting the geometry of the active site and breaking the rules of base-pairing, while at the same time escaping enzymatic mechanisms of error discrimination which scan for the correct geometry of the minor groove.

In contrast, by showing that the *A*-form greatly attenuates the sequence-dependent structural alterations in hotspots, this study suggests that the *A*-conformation of the nascent template-primer duplex at the vicinity of the polymerase active site will contribute to fidelity. The *A*-form may play the role of a structural buffer which preserves the correct geometry of the active site for all sequences. The detailed comparison of the conformation of the nascent template-primer duplex in the available crystal structures of DNA polymerase-DNA complexes shows that polymerase β, the least accurate enzyme, is unique in binding to a *B*-DNA duplex even close to its active site. This model leads to several predictions which are discussed in the light of published experimental data.

Keywords: unusual DNA structure; spontaneous mutagenesis; microsatellite instability

Introduction

The accuracy of DNA replication is fundamental for the genetic stability of the cell. From bacteria to higher eukaryotes, error frequencies are remarkably low, between 10^{-9} and 10^{-10} per base replicated (Echols & Goodman, 1991). These low mutation rates are achieved by multiple steps of error discrimination including base selection by DNA polymerases, 3′-5′ exonucleolytic proofreading and post-replicative DNA repair (Kunkel, 1992). Although the DNA synthesis errors play a role in aging and disease, spontaneous mutations also provide the opportunity for genetic variation and are a primary basis for the evolution. Replication errors occur non-randomly and result from an intricate interplay of intrinsic polymerase fidelity and DNA sequence effects.

A principal determinant of polymerase fidelity is the geometric selection of nucleotide for insertion into DNA (Echols & Goodman, 1991; Goodman, 1997). Crystallographic studies of oligonucleotide duplexes containing mismatched base-pairs have shown that the geometric equivalence and the pseudo 2-fold symmetry of a correct Watson-Crick base-pair is violated in mismatched bases, thus providing a geometrical means for selection by the

E-mail address of the corresponding author: timsit@ibpc.fr

enzymes (Kennard, 1987). Indeed, recent structural studies on DNA polymerases have shown that the accurate discrimination of a correct base-pair is achieved by the tight steric complementarity between a Watson-Crick base-pair and the polymerase active site. Hydrogen-bonding interactions which probe the correct geometry of the base-pairs within the minor groove of the nascent template-primer duplex also participate in error discrimination (Kunkel & Wilson, 1998; Brautingam & Steitz, 1998; Beard & Wilson, 1998).

The role of DNA sequence and structure in polymerase fidelity is less well understood. The analysis of the DNA sequence surrounding the mutational hotspots has led to several models of sequence-directed mutagenesis (Ripley, 1990; Kunkel, 1990). Primer-template misalignment has been proposed to explain the high rates of spontaneous frameshift or substitution mutations in homopolymeric runs (Streisinger *et al.*, 1966; Kunkel, 1990). Some studies have suggested that alterations of base stacking (Petruska & Goodman, 1995; Mendeman *et al.*, 1989), unusual backbone flexibility (Blake *et al.*, 1992; Mitra *et al.*, 1993; El Antri *et al.*, 1993; Hess *et al.*, 1994) or alternative DNA structures (Freund *et al.*, 1989) can affect the accuracy of replication at hotspots. However, little is known about the contribution of the sequence-dependent structural variability of DNA in the initiation of replication errors, and the molecular mechanisms by which DNA structure affect the accuracy of DNA replication are still unknown.

Several years ago a crystallographic study suggested that the shift in base-pairing observed in the major groove of $(CA)_n$ tracts could participate in the initiation of frameshift errors in hotspots (Timsit *et al.*, 1989, 1991). Indeed, in both prokaryotic and eukaryotic cells, very high rates of spontaneous frameshift mutations at $(CA)_n$ tracts were found to result from polymerase slippage (Levinson & Gutman, 1987; Strand *et al.*, 1993). The high rate of slippage at CA repeats which is revealed in cells deficient in post-replicative mismatch repair is responsible for microsatellite instability and is associated with cancer and other human diseases (reviewed by Kunkel, 1993; Loeb, 1994; Karran, 1996). A better understanding of the molecular basis of sequence-directed mutagenesis requires an answer to the following questions: (1) are there predictible structural features of DNA which could account for higher error rates? (2) What are the molecular mechanisms which relate DNA structure to the polymerase fidelity?

Here, the analysis of the crystal structures of mutational hotspots including $(CA)_n$ and $(A)_n$ tracts (with $n \geqslant 2$) in different intermolecular contexts reveals that, in the *B*-form DNA, these sequences share common structural alterations in base-pairing and stacking which may explain the high rate of replication errors. In contrast, the *A*-form greatly attenuates sequence-dependent structural alterations. A molecular mechanism providing a structural basis for sequence-directed mutagenesis is proposed in the light of the recent structural studies on DNA polymerases. The model involves polymerase β (Pelletier *et al.*, 1994) bound to a misaligned double helix and suggests that ''Janus-like'' structural features of the nascent template-primer duplex can mislead the nucleotide incorporation while escaping from the error discrimination mechanisms used by the polymerases. The model also suggests that the *A*-conformation of the duplex observed at the vicinity of the active site in many polymerases contributes to polymerase fidelity by attenuating the sequence-dependent alterations. A detailed comparison of the conformation of the nascent template-primer duplex in the available crystal structures of DNA polymerase-DNA complexes indeed shows that polymerase β, the least accurate polymerase, is unique in binding to a *B*-DNA duplex.

Results and Discussion

Common features in the *B*-DNA helical structure of frameshift mutational hotspots: comparison of $(CA)_n$ and $(A)_n$ tracts in different intermolecular contexts

Crystallographic, NMR and biochemical studies have revealed that CA steps, CAC triplets and $(CA)_n$ tracts exhibit unusual structural and dynamic features such as shifted base-pairing, kinking and a higher rate of base-pair opening (reviewed by Timsit & Moras, 1996, and references cited therein). For example, *B*-DNA helices containing CA repeats observed in the crystal structures of DNA duplexes or protein-DNA complexes exhibit an irregular geometry with marked alterations in base stacking, base-pairing and in the sugar-phosphate backbone conformation (Figure 1(a) and (b)). In the tet dodecamer (Timsit *et al.*, 1991), the bases are not paired with their Watson-Crick complements but with their direct 5′ neighbors, on the opposite strand (Figure 1(a)). The bases form a set of consecutive A·G and C·T mispairs on the major groove side, while the base-pairing remains unaltered in the minor groove. A magnesium cation stabilizes the unfavorable approach of the shifted mismatch between the G7 and the G17 bases by bridging their two O6 carbonyl groups (see Figure 2 by Timsit & Moras, 1996). This geometry is achieved by a combination of high propeller twists, opening angles, stagger and rise (Table 1). The unusual high values of the stagger at each step indicate that one strand has moved in the 5′-3′ direction relative to the other. The duplex may be therefore considered as a pre-slipped double helix. The C9-A10 and T15-G16 steps are characterized by a severe distorsion of the sugar-phosphate backbone, including an unusualy close approach of consecutive phosphate groups, with phosphorus-phosphorus distances of 6.2 and 5.7 Å, respectively. Figure 1(b) shows that the CA repeats exhibit a similar irregular conformation in the CAP-DNA co-crystal structure (CAPS1) (Schultz

Table 1. Structural features of $(CA)_n$ and $(A)_n$ tracts in different molecular contexts

		Stagger (Å)	Buckle (deg.)	Propeller twist (deg.)	Opening (deg.)	Rise (Å)	Slide (Å)
tet	**C9-G16**	1.8	−6.9	−27.8	20.8	3.0	1.3
$(CA)_2$	**A10-T15**	1.9	−2.7	−21.0	30.9	3.9	−1.2
	C11-G14	2.0	−13.6	−25.5	−0.1	3.0	−0.8
	A12-T13	1.3	−10.5	−10.7	10.5		
CAPS1a	**C5-G27**	−1.0	−18.0	−32.0	16.6	4.3	1.5
(chain D)	**A6-T26**	−0.3	−10.3	−28.2	1.0	3.4	−0.2
$(CA)_2C$	**C7-T25**	−0.3	−7.7	−16.4	2.6	2.9	0.5
	A8-T24	0.7	−32.2	−24.2	5.8	4.3	−1.7
	C9-G23	0.5	−5.6	−21.0	1.2		
CAPB3b	**T20-A20**	−0.2	−32.0	−40.0	4.4	3.0	−0.9
$(A)_5$	**T21-A21**	0.0	−35.3	−48.7	−5.2	2.7	−0.4
	T22-A22	−1.9	−14.0	−47.8	5.8	3.5	−1.6
	T23-A23	0.1	−9.3	4.1	−17.0	2.5	−0.9
	T24-A24	−0.5	6.5	−18.4	−9.6		
RAP Ib	**C10-G30**	0.0	7.2	−3.6	1.7	3.6	0.9
(ch. C, D)	**A11-T29**	0.1	−1.3	−23.2	1.3	3.4	−0.4
$(CA)_3C$	**C12-G28**	−0.2	−1.9	−3.4	4.4	3.1	0.9
	A13-T27	−0.1	−11.3	−13.9	9.7	3.5	−1.0
	C14-G26	−0.2	5.2	−10.1	−1.6	4.0	−1.3
	A15-T25	−0.74	−10.3	−5.0	8.7	3.2	−0.8
	C16-G24	0.7	7.6	−3.2	−1.4		
Atetr	**A5-T4**	0.4	3.1	−14.4	0.7	3.0	−1.5
$(AC)_2$	**C6-G3**	0.0	5.9	−17.7	4.4	2.9	−1.3
	A7-T2	0.1	2.5	−13.6	6.4	2.7	−1.8
	C8-G1	−0.2	11.1	−13.0	1.9		

The helical parameters were calculated with Curves (Lavery & Sklenar, 1989) on the DNA coordinates of the $(CA)_n$ tracts of the dodecamer d(ACCGGCGCCACA) (tet) (Timsit *et al.*, 1991), the CAP-DNA complex (CAPS1) (Schultz *et al.*, 1991), the RAP1-DNA complex (RAP) (König *et al.*, 1996), the $(AC)_2$ tract of the tetragonal form of the octamer d(GTGTACAC) (Atetr) (Jain *et al.*, 1989) and the $(A)_5$ tract of the CAP-DNA complex (CAPB3b) (Parkinson *et al.*, 1996b) which are displayed in Figure 1.

et al., 1991). The *B*-DNA duplex is again characterized by a pronounced unstacking of the bases with very high buckle, rise and propeller twist angles (Tables 1 and 2) leading to the formation of shifted base-pairs and bifurcated hydrogen bonds between crossed Watson-Crick donor and acceptor groups in the major groove. Similar to the tet structure, the bases interact with the 5′ neighbors of their complements (Figure 2(b)) but here, the alterations do not propagate along the whole helix.

$(CA)_n$ tracts and $(A)_n$ tracts share common structural features. In most of DNA crystal structures, the $(A)_n$ tracts adopt a unique geometry in which the consecutive adenine bases form an array of bifurcated hydrogen bonds with the 5′-neighbors of their complements (Coll *et al.*, 1987; Nelson *et al.*, 1987; Aymami *et al.*, 1989). The crossed Watson-Crick interactions are produced by very high negative propeller twist angles, negative inclination and unusual sugar puckers (DiGabriele, 1993; Brahms *et al.*, 1992). For example, the alterations found in the $(A)_n$ tracts of the crystal structure of the phage 434 repressor-DNA complex (Aggarwal *et al.*, 1988) or the CAP-DNA complex (CAPB3) (Figure 1(c) and Table 1) (Parkinson *et al.*, 1996b) are reminiscent of that of tet. This analysis shows that, within the *B*-DNA family, the DNA sequences in which one strand is Watson-Crick donor (N6 or N4 amino group) and the other one is Watson-Crick acceptor (O6 or O4 carbonyl group) viewing the major groove can form altered DNA structures, with very high propeller twist and bifurcated or shifted base-pairs in the 5′-3′ direction.

A-DNA attenuates sequence-dependent structural variations

Figure 1(d) shows that in the crystal structure of the RAP1-DNA complex (RAP), the $(CA)_n$ tracts adopt a more regular structure (König *et al.*, 1996). The base-pairs are stacked in parallel with a low propeller twist (Table 1). This finding is somewhat surprising, since $KMnO_4$ reactivity studies in the RAP1-recognition sequence anticipated aberant base stacking and pairing upon protein binding (Gilson *et al.*, 1993). The alternance of high and low individual twists at CA and AC steps, respectively (see Table 1 by König *et al.*, 1996), seems to contribute to maintain the planarity of the base-pairs in building a base stacking pattern which disfavors propeller twisting. Similarly, the high twists (50°) at the CA/TG steps in monoclinic decamer duplexes have been associated with a low propeller twist and rise (Privé *et al.*, 1991). However, another parameter contributing to the uniformity of the base stacking pattern is the *A*-character of the double helix, as indicated by the high negative slide (local) and X-displacement observed in many steps (Tables 1 and 2).

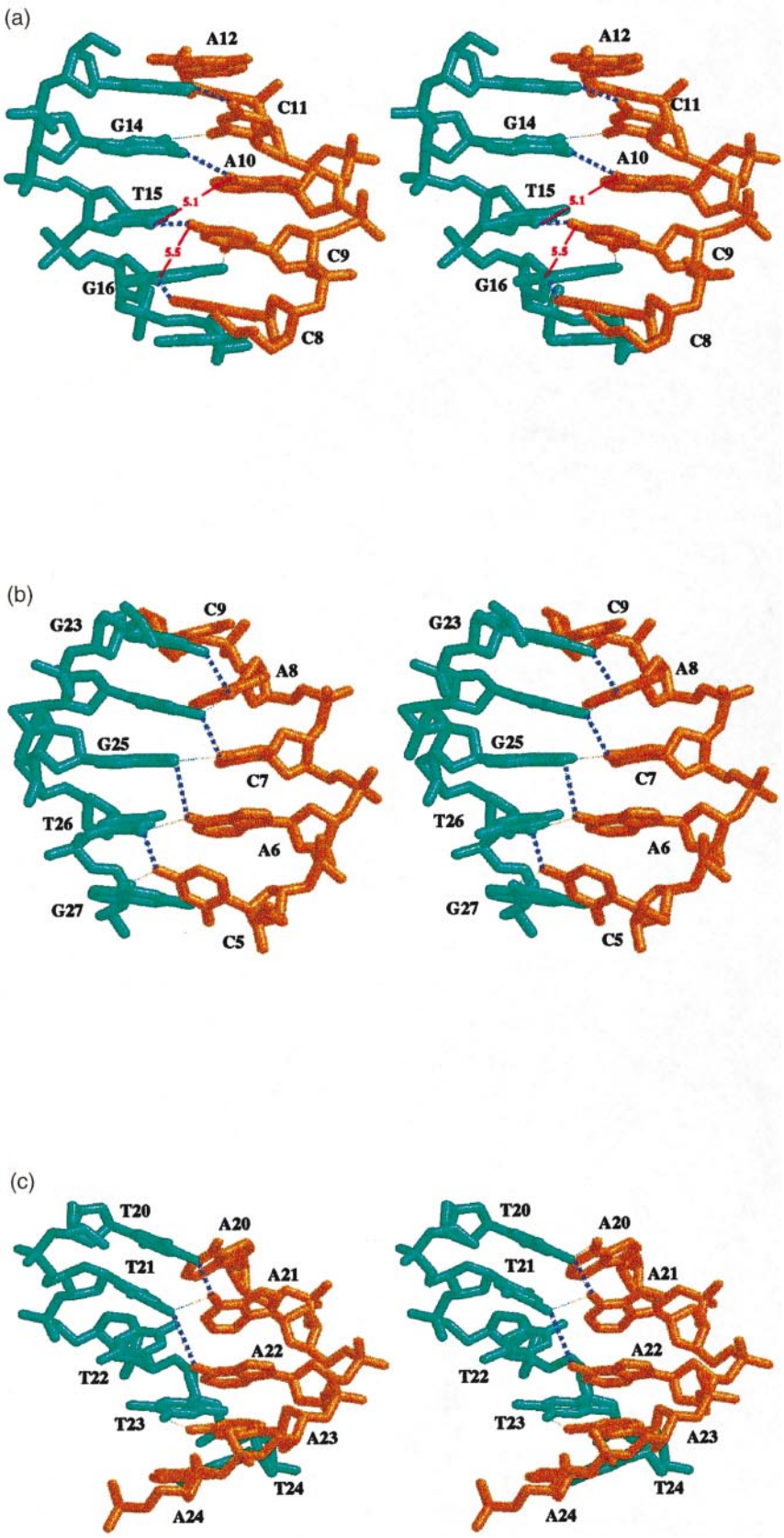

Figure 1 (*legend opposite*)

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.